# ACCURATE PROJECTION OF SMALL-SCALE RASTER DATASETS

**Usery, E.L., Seong, J.C., Steinwand, D.R. and Finn, M.P.**

U.S. Geological Survey, Mid-Continent Mapping Center, Rolla, Missouri.
Northern Michigan University, Department of Geography, Marquette, Michigan.
Science Applications International Corporation, EROS Data Center, Sioux Falls, South Dakota.
Tel: 573-308-3837. Fax: 573-308-3652. E-mail: usery@usgs.gov

**ABSTRACT**

It has become apparent that the validity of raster data transformations for various map projections can have significant effects on Earth modeling results. This research investigated three related geospatial information science topics: dynamic projections; the relationships between representations and features, categories, and patterns; and resampling procedures. Empirical investigation compared results of projecting global datasets from geographic coordinates to ten common projections. Analysis of categorical areas indicates a dependence on resolution with errors among equal-area projections ranging between 2 and 20 percent. Error analysis indicates problems in reprojection with specific projection-to-projection transformations yielding better results. Examination of the resampling methods associated with projection of categorical datasets has resulted in the development of a new method. Integrating these lessons, we developed a decision support system for selecting an optimum projection.

## 1. INTRODUCTION

Modeling regional, continental, and global changes of the Earth has become a critical need in recent years because of the impact of human activities on climatic and biotic systems. As high-resolution global data sets in raster format have become more readily available and the computational power for modeling has improved, it has become apparent that the integrity and validity of data transformations can have significant effects on modeling results. Through empirical analysis with actual datasets, error analysis of pixel gain and loss, and development of new resampling procedures, we have investigated the effects of projection transformations and resampling in a global context.

Our empirical work has compared results of projecting global vegetation, ecological zone, temperature, rainfall, elevation, population, and land cover data from geographic coordinates (angular grids based on latitude and longitude) to ten common projections. The global raster datasets occur at resolutions from 30-arcsec to 1-degree cells. Analysis of areas of specific categories indicate a dependence on resolution with errors among equal-area projections of 20 percent for 1-degree cells, 10 percent for one-half degree cells, and only one to two percent for 30-arcsec cells. Thus, as raster resolution increases the projection transformations designed for point data, become more accurate.

Error analysis also indicates problems in reprojection with specific projection-to-projection transformations yielding better results. A common data representation of processed combinations of several acquisitions of small-scale raster image datasets is the composite pixel that is a representative value of a particular ground location over a period of time (Eidenshink and Faundeen, 1994). Examination of the resampling methods associated with these composite pixels and their associated projection as categorical datasets has resulted in the development of a new method, which offers user selection options in the resampling approach and yields superior visual results.

This research investigated three goals theoretically and empirically. In addition it employed the developed empirical base of knowledge to develop a prototype expert system for map projection of small-scale raster data for regional, continental, and global dataset modeling.

The three theoretical goals were:
- The development of a dynamic projection, which adjusts projection formulas for latitude based on raster cell size to maintain equal-sized cells.
- The investigation of the relationships between the raster representation and the distortion of features, number of categories, and spatial pattern.
- The development of an error correction and resampling procedure based on error analysis of raster projection.

Proceedings of the 21st International Cartographic Conference (ICC) Durban, South Africa, 10 – 16 August 2003
*'Cartographic Renaissance'* Hosted by The International Cartographic Association (ICA)
ISBN: 0-958-46093-0 Produced by: Document Transformation Technologies

From this knowledge base, we developed a prototype decision support system (DSS) for selecting an optimum projection considering various factors such as pixel size, areal extent, number of categories, spatial pattern of categories, resampling methods, and error correction methods.

In investigating a dynamic projection, we assumed that modelers could accurately project regional and global raster data with appropriate equations that account for raster cell size and latitudinal position. In researching the relationships between the raster representation and the distortion of features, number of categories, and spatial pattern, we imagined that scale factors explain the impact of distortion on representation, and, further, that more categories and more complex spatial patterns cause more errors. Additionally, we hypothesized that error correction and resampling methods or procedures can be used for optimizing projection accuracy of small-scale raster datasets. Applying this knowledge, we then developed the DSS for selecting an optimum projection. These results (the equations, models, methods, and DSS) have potentially positive applications and impacts on many U. S. national and international programs involving the use of regional, continental, and global raster data such as the Global Land Cover Project and climate change research (Usery *et al*., 2001; Usery *et al*., 2003; Steinwand, 2003; Seong, 2003).

## 2. DYNAMIC PROJECTION

In order to assess the feasibility of creating a dynamic projection to preserve global areas in a raster representation, we first established a control standard to judge results of empirical map projection transformations. We established this control by developing global raster datasets in spherical coordinates with resolutions of one degree, 30 minutes, and 30 arc seconds. The area of each cell in these datasets was then computed through a numerical integration procedure (Usery *et al*., 2003). The result is a dataset in spherical coordinates where each cell contains a digital number value equivalent to the area of the cell. These cell values represent truth with respect to the actual ground area. Table 1 presents some sample values of these single cell areas at various latitudes (Usery *et al*., 2003). To quantify the accuracy of the area values of an empirical, thematic global dataset (such as land cover), we calculated the total area for occurrences of each land cover type thus yielding exact areas for each category. These summed values became the basis for comparison with projected data of the same type.

Table 1. Areas of Single Pixels Computed from Spherical Coordinates (in Meters$^2$) (from Usery *et al*., 2003).

| | | **Pixel Size** | |
|---|---|---|---|
| **Latitude** | **30 ’’** | **0.5 º** | **1 º** |
| **0** | 858,631 | 3,091,035,692 | 12,363,671,878 |
| **15** | 829,390 | 2,982,220,448 | 11,970,315,668 |
| **30** | 743,628 | 2,670,171,821 | 10,761,202,175 |
| **45** | 607,188 | 2,176,155,408 | 8,818,730,582 |
| **60** | 429,370 | 1,533,837,609 | 6,275,272,108 |
| **75** | 222,291 | 786,991,318 | 3,304,173,896 |
| **90** | 63 | 13,487,417 | 107,896,706 |

We can project each raster line in spherical coordinates to a corresponding raster line in plane coordinates allowing us to maintain the exact area from spherical coordinates in the plane representation. However, this procedure prohibits the joining of adjacent raster lines because only pixels in a single raster line in plane space have the same areas. Therefore, only single raster lines can be displayed. The logical completion of this approach is to resample all projected lines to achieve a common pixel size.

## 3. ERROR THEORY

The scale factor model shows promise in modeling error resulting from replicated and lost pixel values in the projection transformation process. A theoretical examination revealed that error results in two forms: areal size change of pixels and categorical error resulting from loss or duplication of pixels. We developed a scale factor model to provide a computation of the resulting error from specific projections. The model was experimentally tested with the Equal-Area Cylindrical, Sinusoidal, and Mollweide projections. Results indicate that the Sinusoidal projection is subject to smaller errors in projecting raster data than the other projections tested (Seong and Usery, 2001).

In addition, we applied the scale factor model in image reprojection and observed a significant change in pixel values. Six possible reprojections among the Equal-Area Cylindrical, the Mollweide, and the Sinusoidal were tested. Results show that reprojection accuracy can be explained using the ratios of scale factor changes along vertical and horizontal axes between source and target projections. We modeled the reprojection accuracy as the reciprocal of the maximum scale factor change along either the vertical or the horizontal axis:

Reprojection Accuracy = 1 / Maximum ( X2 / X1, Y2 / Y1)

where, 1 and 2 represent the original and target projections respectively, X is the horizontal scale factor, and Y is the vertical scale factor.

The model explains reprojection accuracy very well. The model accuracy is very sensitive to the skew effect that leads to significant error increase (Seong, 2003)

Because the Sinusoidal projection appeared most accurate, it was investigated in more detail. The Sinusoidal projection showed 99.50% and 98.35% categorical accuracies when 54 sample data sets were reprojected from the UTM to the Sinusoidal and from the Sinusoidal to the UTM, respectively. Table 2 shows the results of reprojection accuracies between global projections and UTM, a very accurate local projection.

Table 2. Reprojection accuracies between global and local projections.

| Original Projection | Target Projection | Minimum Accuracy (%) | Maximum Accuracy (%) | Average Accuracy (%) |
|---|---|---|---|---|
| UTM | Sinusoidal | 98.5 | 99.9 | 99.5 |
| UTM | Mollweide | 48.2 | 99.6 | 85.8 |
| UTM | Eckert IV | 19.7 | 99.8 | 76.9 |
| UTM | Hammer-Aitoff | 53.2 | 100.0 | 87.0 |
| UTM | Geographic | 100.0 | 100.0 | 100.0 |
| Sinusoidal | UTM | 90.5 | 100.0 | 98.4 |
| Mollweide | UTM | 68.0 | 99.8 | 89.5 |
| Eckert IV | UTM | 29.9 | 99.3 | 85.2 |
| Hammer-Aitoff | UTM | 59.9 | 100.0 | 87.2 |
| Geographic | UTM | 9.2 | 99.9 | 65.3 |

In addition, we investigated the effect of skewness and number of categories with sample data sets. Figure 1 shows the extent of increase of raster representation error due to skew effect.

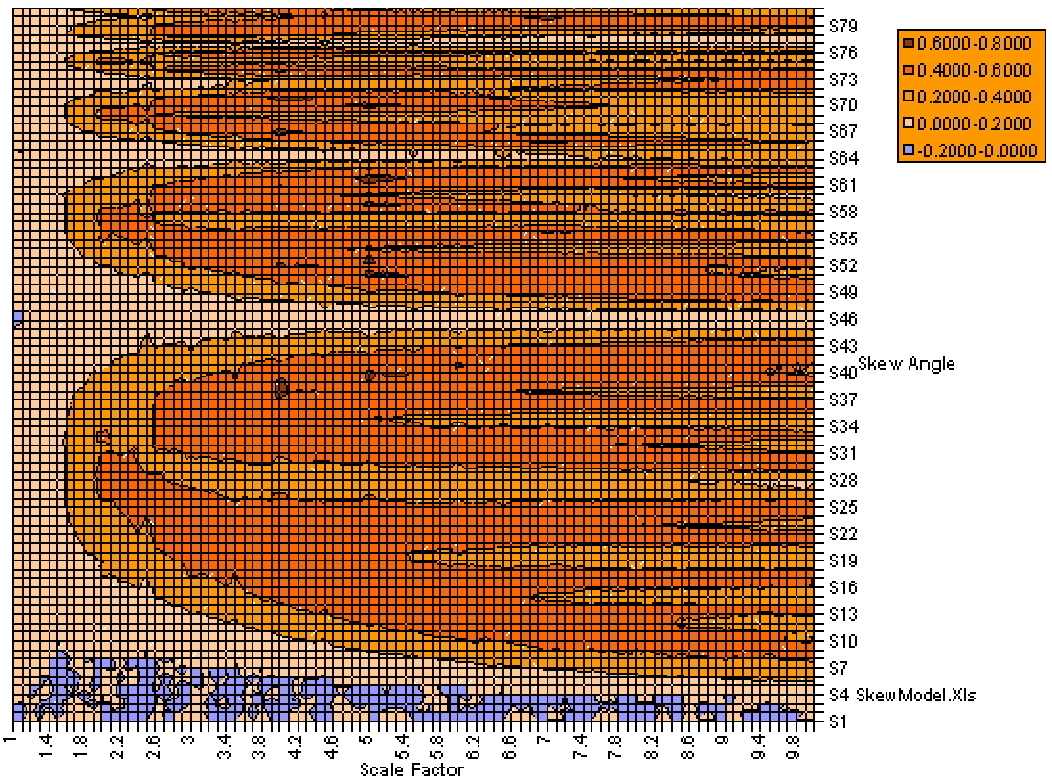


Figure 1. Accuracy increase due to the change of skew angle.

The Y-axis in Figure 1 indicates the skew angles in degrees. The graph values indicate the ratio of pixel duplication in percents. For example, with the scale factor of 4.2 and skew angle of 22 degrees, we will expect 40% - 60% increase in categorical accuracy. Therefore, if a place has a modeled categorical accuracy of 50% and the scale factor and skew angle are 4.2 and 22 degrees respectively, we will expect an increased accuracy between 70% (=50% + 50% * 40%) and 80% (=50% + 50% * 60%).

## 4. RESAMPLING

Commonly available techniques for resampling categorical data typically consist of nearest neighbor-like resampling methods. These methods are chosen because the alternatives, interpolating methods, do not maintain categorical values. Indeed, commercial applications programmers have designed most of the software tools available today for single scene, signal-based remote sensing image data, where the image extent usually extends only a few hundred kilometers. However, not many existing tools are designed for small-scale datasets (Steinwand, 2003).

The typical nearest neighbor algorithm for categorical resampling takes the center of a pixel as point data and reprojects that coordinate into the new projection space. Because the resulting coordinate is often not at an exact pixel location, it is rounded to the nearest pixel position and that pixel's value is used to populate the resulting output image. This method often results in imagery that is not representative of the original image due to varying amounts of geometric distortion present in the projection transformation.

We developed a new method for resampling categorical data. This new resampling method treats pixels not as points, but as areas. In summary, the new algorithm maps the four corner coordinates of the image pixels between the two projections and determines the number of input image pixels that go into making the resulting output image pixel. If one pixel goes into the making of the output image pixel, the nearest-neighbor approach is used. More often, multiple input image pixels go into one output image pixel. In this case, simple statistical methods are used to determine the output pixel value. This many-to-one condition is present in transformations of similar pixel sizes where geometric distortions are great. The algorithm keeps track of pixel values that were present in the input area, but not used. It then writes them to a statistical image that can assist the user in determining the extent of data loss or misrepresentation (Steinwand, 2003).

The two images below illustrate the output of the algorithm. Figure 2 was processed using the nearest neighbor resampling method and appears noisy. Figure 3 was processed with the maximum occurring pixel method and appears smoother. Both figures 2 and 3 incorporate a reprojection to Mollweide and a downsampling to 50-km pixels from the original 1-km data (Steinwand, 2003).

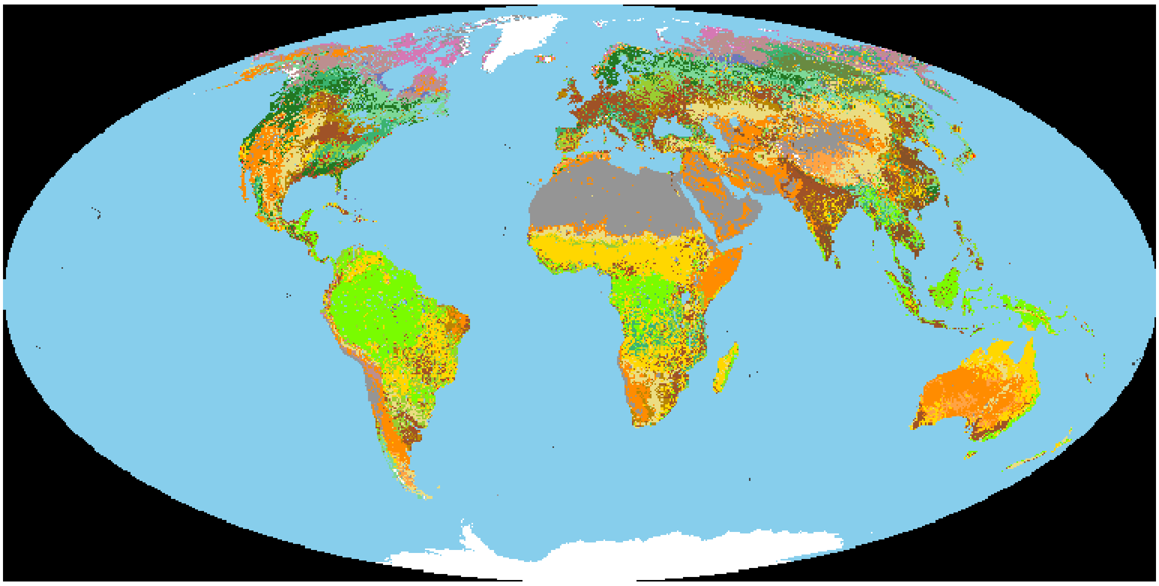

Figure 2. Extreme downsampling and reprojection with nearest neighbor.

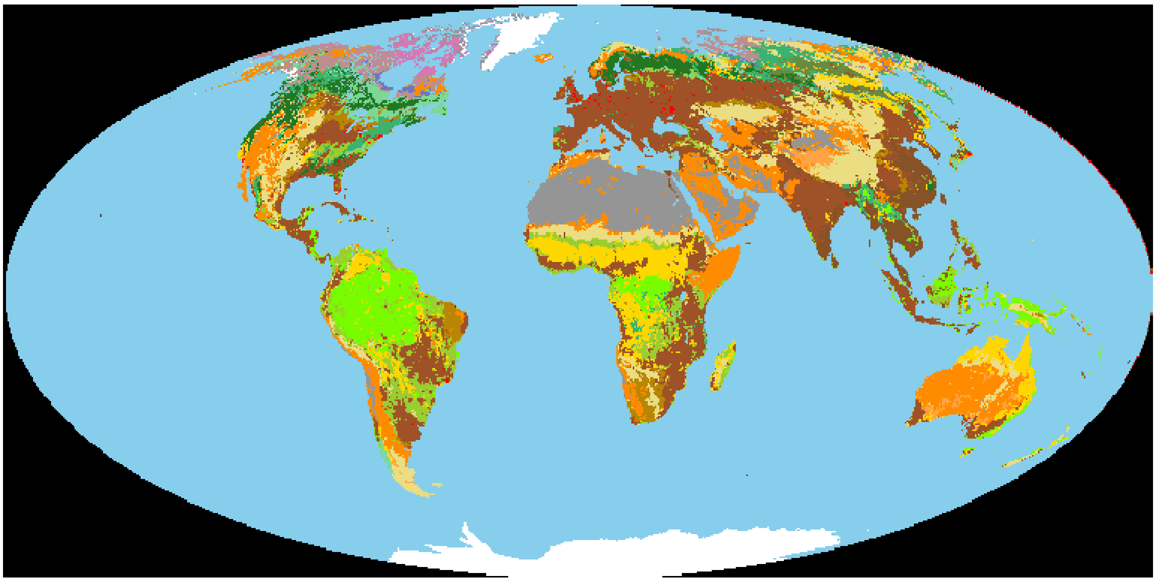

Figure 3. Extreme downsampling and reprojection with the new algorithm.

## 5. DECISION SUPPORT SYSTEM (DSS) FOR MAP PROJECTIONS

We implemented a prototype of a web-based system for supporting map projection selection. This prototype DSS is for projections of small-scale data sets, such as those with regional, continental, or global geospatial extent. The purpose is to assist users in selecting a projection when using a commercial GIS software package based on the characteristics of their data and their intended uses. The prototype focuses on user input of geographic area and geometric characteristic preservation to drive the suggested projection or projections (Figure 4).

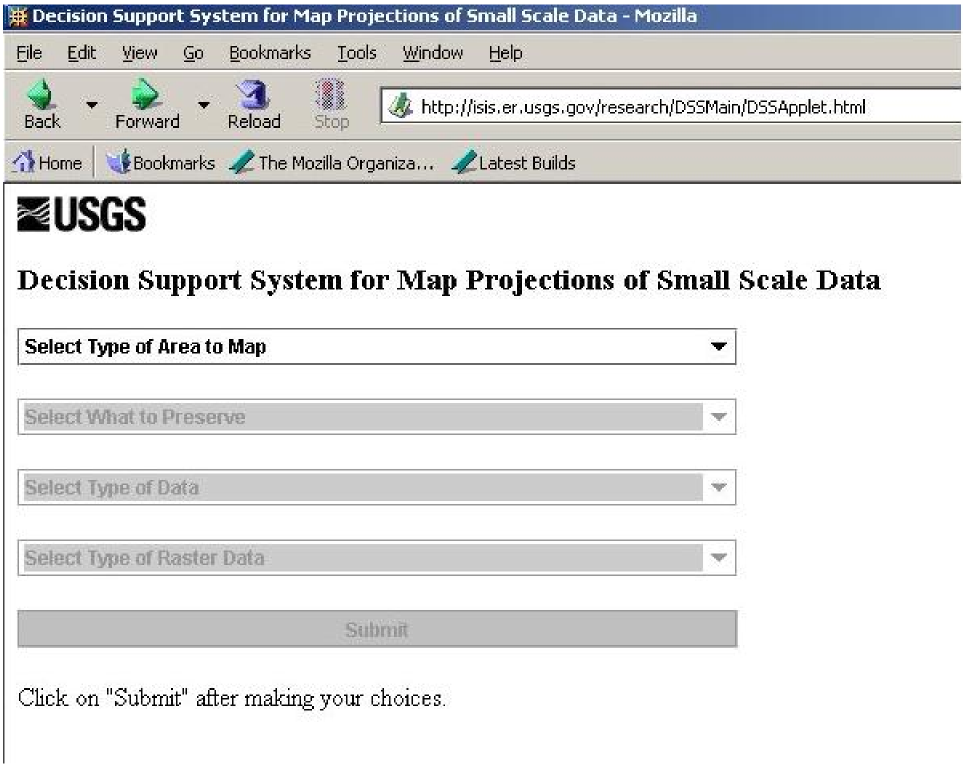


Figure 4. Screen shot from the prototype DSS for map projection.

The current prototype was developed primarily using the Java[1] programming language. This allows a user to execute the DSS interactively across the web independent of their hardware platform. It employs a system of drop down menus

[1] Java is a trademark of Sun Microsystems, Inc

for selecting the parameters or characteristics of the projections. Figure 5 shows an example of one of the drop down menus. We plan to add a tutorial on map projections for the final development.

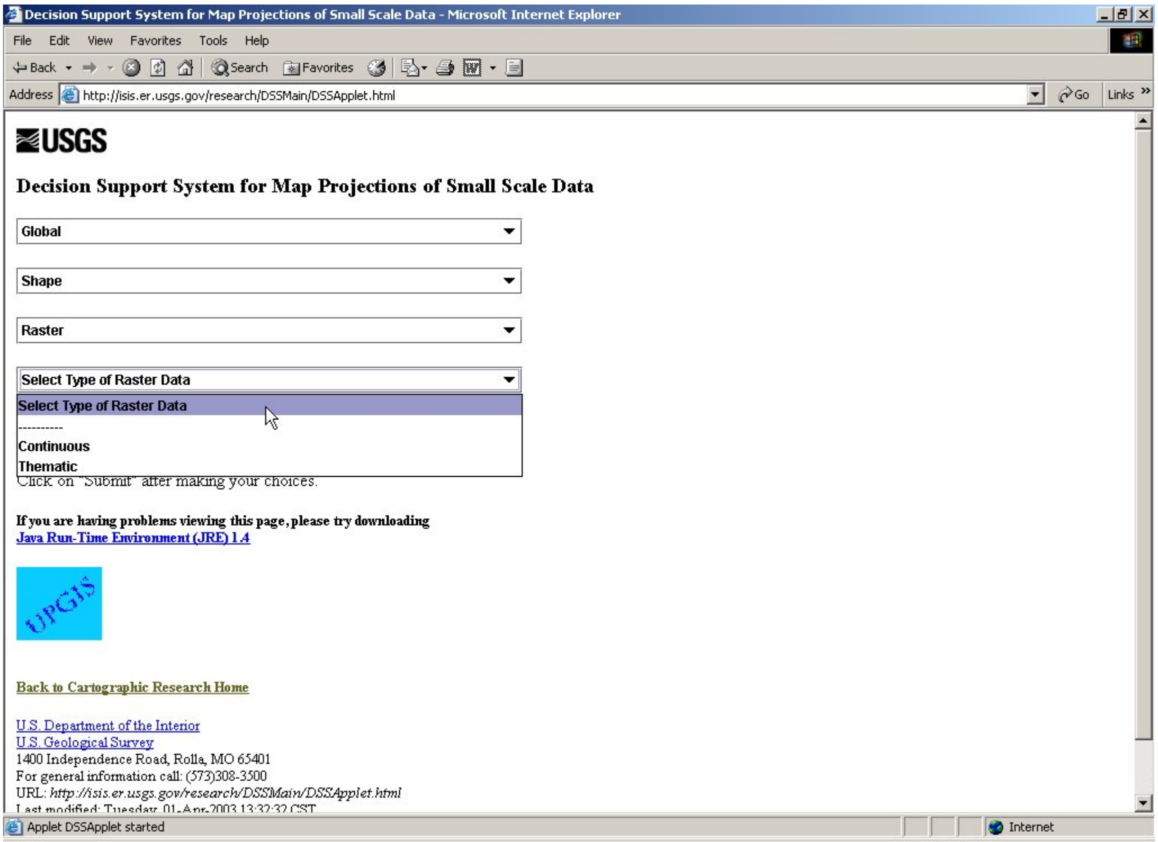


Figure 5. Prototype DSS -- an example of a drop down menu.

## 6. CONCLUSIONS

This research demonstrates procedures to accurately project small-scale raster data with equations that account for cell size and latitudinal position and presented a scale factors that explained the impact of distortion on representation. It appears promising that the empirical results confirm the theoretical results from previous work (for example, Kimerling, 2002). In addition, it demonstrated a new method of categorical resampling to optimize projection accuracy of small-scale raster datasets. This method takes into account pixel areas instead of merely a point within a given pixel. It then maps the footprint of the output pixel back into the input image, taking into consideration all pixels under that footprint, rather than just the nearest neighbor. Collectively, these procedures and methods drove the resulting DSS for selecting a map projection. Future applications should examine the feasibility of designing a system for global raster representation and analysis.

## 7. REFERENCES

[1] Eidenshink, J.C., and Faundeen, J.L., 1994, The 1-km AVHRR global land data set: First stages in implementation. *International Journal of Remote Sensing*, Taylor and Francis, Ltd., v. 15, no. 17, pp. 3443-3462.

[2] Kimerling, A.J., 2002, Predicting data loss and duplication when resampling from equal-area grids. *Cartography and Geographic Information Science*, v. 29, no. 2, pp 111-126. ACSM, Gaithersberg, MD.

[3] Seong, J.C., 2003, Modeling the accuracy of image data reprojection. Submitted to *International Journal of Remote Sensing,* Taylor and Francis, Ltd.,

[4] Seong, J.C., and Usery, E.L., 2001, Modeling raster representation accuracy using a scale factor model. *Photogrammetric Engineering and Remote Sensing*, v. 67, no. 10, pp. 1185-1191.

[5] Steinwand, D.R., 2003, A new approach to categorical resampling. *Proceedings of* the *American Congress on Surveying and Mapping Spring Conference*, Phoenix, AZ. ACSM, Gaithersberg, MD.

[6] Usery, E.L., Finn, M.P., Cox, J.D., Beard, T., Ruhl, S. and Bearden, M., 2003, Projecting global datasets to achieve equal areas. *Cartography and Geographic Information Science*, v. 30, no. 1. ACSM, Washington.

[7] Usery, E.L., Seong, J.C., Steinwand D.R., and Finn, M.P., 2001, Methods to achieve accurate projection of regional and global raster databases. *U.S. Geological Survey Open-File Report 01-383*.

# ACCURATE PROJECTION OF SMALL-SCALE RASTER DATASETS


**Usery, E.L., Seong, J.C., Steinwand, D.R. and Finn, M.P.**

U.S. Geological Survey, Mid-Continent Mapping Center, Rolla, Missouri.
Northern Michigan University, Department of Geography, Marquette, Michigan.
Science Applications International Corporation, EROS Data Center, Sioux Falls, South Dakota.
Tel: 573-308-3837. Fax: 573-308-3652. E-mail: usery@usgs.gov


**Biography**

Michael P. Finn holds a BS in Geography with a Minor in Cartography and Map Technology from Southwest Missouri State University and an MS in Civil Engineering from Virginia Polytechnic Institute & State University. He has worked as a Computer Programmer with the US Geological Survey for 4 years and has 17 years with the US Department of Defense; 10 years with the US Air Force and 7 years with the Defense Mapping Agency. He is Past President, Central Region, American Society for Photogrammetry and Remote Sensing. His research interests are in quantitative approaches to imaging in environmental modeling and GIS; in geodesy and spatial coordinate systems; and in discrete mathematical and scientific applications for digital geospatial data.